Australian Government
Department of Defence
Defence Science and
Technology Organisation

Evaluation of the 'Mentor' Assessment and Feedback System for Air Battle Management Team Training

Christopher Best and Eleanore Burchat

Air Operations Division
Defence Science and Technology Organisation

DSTO-TR-1942

ABSTRACT

The Mentor software package (Calytrix Technologies, Perth, Western Australia) is gaining popularity within the Australian Defence Force (ADF) as a means by which to manage training objectives, collect performance data and provide feedback for collective training. While the Navy has led the way in the application of this tool, it is now being put forward as an important component of an Air Warfare Assessment and Readiness Evaluation System (AWARES) for the RAAF as well as being included in the suite of tools to be used for exercises involving the Joint Combined Training Centre (JCTC). This report contains an account of an evaluation of the Mentor system and its use to provide performance assessment and feedback during a RAAF Air Battle Management team training event.

Published by

Air Operations Division
DSTO Defence Science and Technology Organisation
PO Box 1500
Edinburgh South Australia 5111   Australia

Telephone: (08) 8259 5555
Fax: (08) 8259 6567

AR-013-795
November 2006

APPROVED FOR PUBLIC RELEASE

Evaluation of the 'Mentor' Assessment and Feedback System for Air Battle Management Team Training

Executive Summary

The Mentor software package (Calytrix Technologies, Perth, Western Australia) is gaining popularity within the Australian Defence Force (ADF) as a means by which to manage training objectives, collect performance data and provide feedback for collective training. While the Navy has led the way in the application of this tool, it is now being put forward as an important component of an Air Warfare Assessment and Readiness Evaluation System (AWARES) for the RAAF as well as being included in the suite of tools to be used for exercises involving the Joint Combined Training Centre (JCTC). Given the widespread interest in this software package and associated training methods within the ADF, it is timely to consider their strengths and potential shortcomings in the context of a thoroughgoing evaluation.

In this report, the use of the Mentor system to provide performance assessment and feedback during collective training events is considered in the context of an Air Battle Management (ABM) command team training exercise and in terms of two standard dimensions of training system evaluation (e.g. Kirkpatrick, 1987); (i) trainee and assessor reactions, and (ii) performance change during the training event. The first dimension - student and assessor reactions to the Mentor training system - was evaluated via qualitative analysis of participants' responses to structured interviews. The second dimension - performance change - was evaluated via quantitative analysis of the Mentor performance ratings obtained during the exercise. When considered together these dimensions speak to a broad spectrum of issues, from user acceptance and perceived strengths and weaknesses, to the potential for the system to contribute to desired changes in student behaviour.

From the outcomes it was clear that all students and instructors involved in this evaluation considered collective training, assessment and feedback to be important activities for improving the effectiveness of RAAF ABM teams. However, they also lamented the fact that the opportunities for collective training come about relatively infrequently when compared to individual training. The evidence presented here suggests that collective training does lead to at least short-term performance improvements on behavioural observation measures related to ABM team tasks and important teamwork dimensions. While the role of the Mentor system in enhancing these improvements was not clear, the system does facilitate planning, assessment and the provision of timely feedback in these contexts and it has broad user acceptance. Clearer evidence regarding the particular effects of the Mentor system will require further investigation in more controlled research environments.

Authors

Dr Christopher Best
Air Operations Division

Dr Christopher Best is a Research Scientist within Air Operations Division's Crew Environments and Training branch (CET). Dr Best holds a B.A.(Hons) in psychology (awarded in 1997) and a PhD in psychology (awarded in 2001). He was a member of the academic staff of the School of Psychology at Deakin University's Melbourne Campus for three years before joining DSTO in 2004. His research interests include human perception, cognition, teamwork, team performance measurement and training.

Eleanore Burchat
Air Operations Division

Eleanore Burchat is a human factors specialist within the Air Operations Division (AOD) of the Defence Science and Technology Organisation (DSTO). She joined DSTO in 2003 after completing an honours degree in psychology. Since joining DSTO, she has worked on a range of issues including the design of digital map displays, the measurement of team performance, team decision making and team training, and research into the human factors relating to UAV flight and to UAV attrition.

Contents

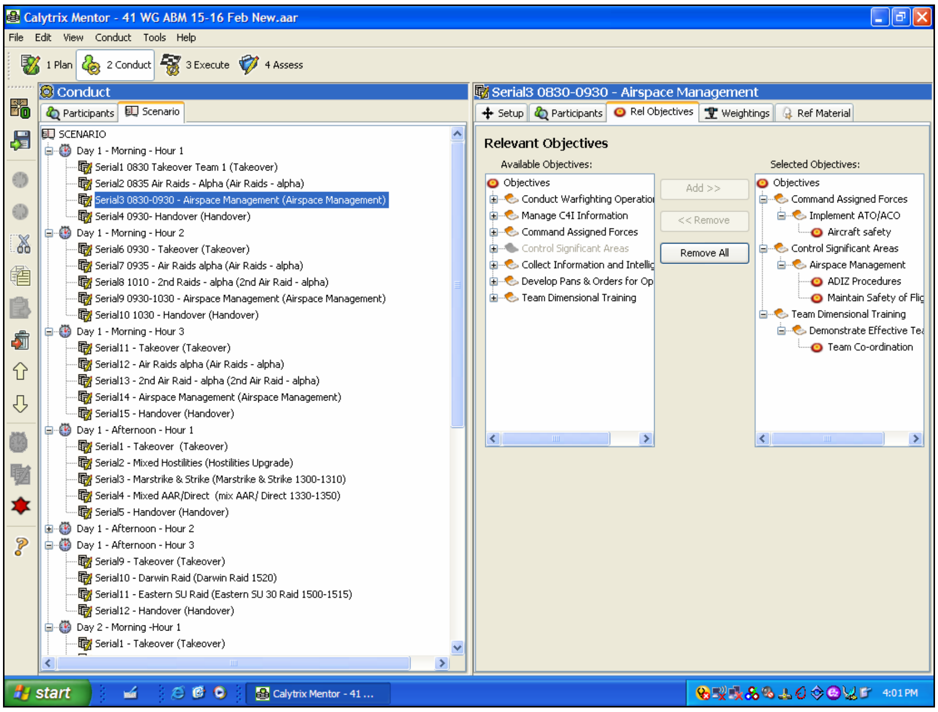
1. Introduction

The Mentor software package (Calytrix Technologies, Perth, Western Australia) is gaining popularity within the Australian Defence Force (ADF) as a means of managing training objectives, collecting performance data and providing feedback for collective training. While the Navy has led the way in the application of this tool, it is now being put forward as an important component of an Air Warfare Assessment and Readiness Evaluation System (AWARES) for the RAAF as well as being included in the suite of tools to be used for exercises involving the Joint Combined Training Centre (JCTC). Given the widespread interest in this software package and associated training methods within the ADF, it is timely to consider their strengths and potential shortcomings in the context of a thoroughgoing evaluation.

The current version of the Mentor system consists of four software tools; (i) the Mentor application itself, (ii) the data entry tool (DET), (iii) the stoplight reports, and (iv) the student handouts. The Mentor main application essentially acts as a database within which users can define trainee and team roles, training objectives and measures, serial events and scenarios composed of those serials. The user can then define relationships between these elements. For example, a team can be defined as being composed of certain roles, each of which has associated training objectives and measures. A scenario can then be assembled from defined serial events, with each event being linked to objectives and measures relevant for each role. A representative screen capture of the main Mentor application is displayed in Panel A of Figure 1. When roles, events, objectives and measures have been defined and linked to create a training scenario, this information can be exported to the DET. The DET essentially acts as an electronic replacement for paper and pencil observer rating sheets. It presents the assessor with an electronic form that can be completed by (i) assigning ratings to measures on a user-defined scale with customisable scoring and verbal scale-point anchors and (ii) providing comments against measures, objectives, and serial events. For this exercise, the DET was presented on a Tablet PC (LG Electronics Model LT20) and comments were recorded via electronic handwriting recognition. A representative screen capture of the DET is displayed in Panel B of Figure 1. Once performance data has been captured via the DET, it can be exported to either or both of two feedback products; the stoplight report and the student handout. The stoplight report presents the assessor’s ratings and comments in a form which can be displayed via a projector in a classroom setting and used to guide after-action review (AAR). The student handout presents the same information in a form which can be printed and given to students so that they can review performance at any time. Examples of the stoplight report and handout are displayed in Panels A and B of Figure 2 respectively.

In this report, the use of the Mentor system to provide objectives management, performance assessment and feedback for collective training are considered in the context of an Air Battle Management (ABM) command team training (CTT) exercise and in terms of two standard dimensions of training system evaluation (e.g. Kirkpatrick, 1987); (i) trainee and assessor reactions, and (ii) performance change during the training event.

A



B

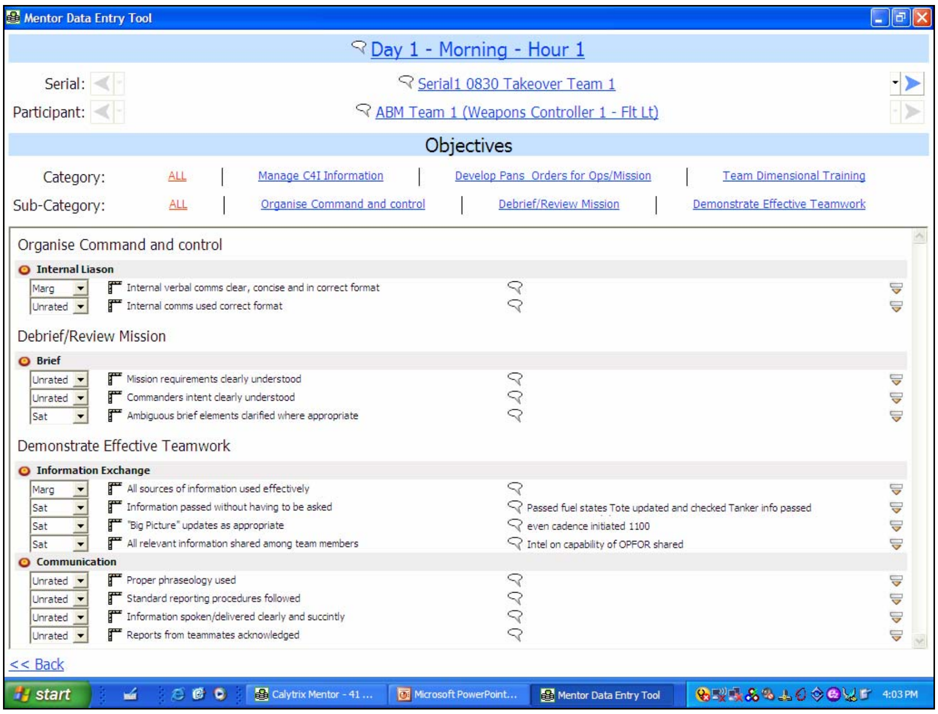
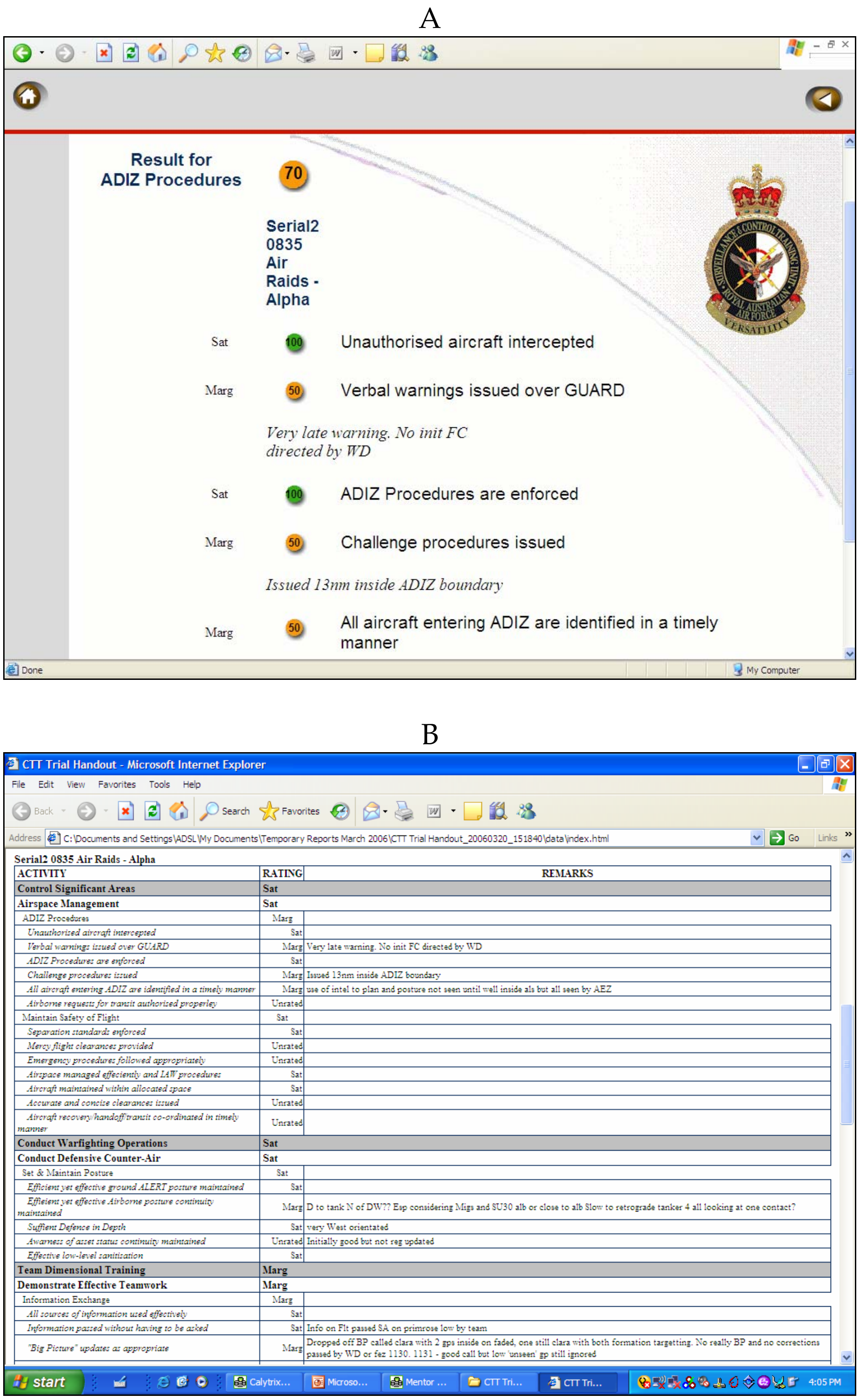


Figure 1. Screen captures from the Mentor software tools. Panel A shows the main Mentor tool and Panel B shows the Data Entry Tool



ACTIVITY	RATING	REMARKS
Control Significant Areas	**Sat**	
Airspace Management	**Sat**	
ADIZ Procedures	Marg	
Unauthorised aircraft intercepted	Sat	
Verbal warnings issued over GUARD	Marg	Very late warning. No init FC directed by WD
ADIZ Procedures are enforced	Sat	
Challenge procedures issued	Marg	Issued 13nm inside ADIZ boundary
All aircraft entering ADIZ are identified in a timely manner	Marg	use of intel to plan and posture not seen until well inside als but all seen by AEZ
Airborne requests for transit authorised properley	Unrated	
Maintain Safety of Flight	Sat	
Separation standards enforced	Sat	
Mercy flight clearances provided	Unrated	
Emergency procedures followed appropriately	Unrated	
Airspace managed effeciently and IAW procedures	Sat	
Aircraft maintained within allocated space	Sat	
Accurate and concise clearances issued	Unrated	
Aircraft recovery/handoff/transit co-ordinated in timely manner	Unrated	
Conduct Warfighting Operations	**Sat**	
Conduct Defensive Counter-Air	**Sat**	
Set & Maintain Posture	Sat	
Efficient yet effective ground ALERT posture maintained	Sat	
Effieient yet effective Airborne posture continuity maintained	Marg	D to tank N of DW?? Esp considering Migs and SU30 alb or close to alb Slow to retrograde tanker 4 all looking at one contact?
Suffient Defence in Depth	Sat	very West orientated
Awarness of asset status continuity maintained	Unrated	Initially good but not reg updated
Effective low-level sanitisation	Sat	
Team Dimensional Training	**Marg**	
Demonstrate Effective Teamwork	**Marg**	
Information Exchange	Marg	
All sources of information used effectively	Sat	
Information passed without having to be asked	Sat	Info on Flt passed SA on primrose low by team
"Big Picture" updates as appropriate	Marg	Dropped off BP called clara with 2 gps inside on faded, one still clara with both formation targetting. No really BP and no corrections passed by WD or fez 1130. 1131 - good call but low 'unseen' gp still ignored

Figure 2. Screen captures from the Mentor software tools. Panel A shows the stoplight report and Panel B shows the student handout

1.1 Evaluation Context

During the week of 13-17 February 2006, human factors researchers from DSTO Air Operations Division conducted an evaluation of the use of the Mentor software package for managing collective training events. The evaluation was performed during an Air Battle Management (ABM) team training exercise held at 41 Wing, RAAF Base Williamtown as part of Surveillance and Control Training Unit's (SACTU) 2006 Fighter Combat Controller (FCC) course. The primary aim of the exercise was to train and evaluate the performance of students in the role of weapons director (WD) of an ABM team (i.e. in the role of team leader). The evaluation of the Mentor software reported here was conducted in parallel with the main training and assessment effort. The DSTO team collaborated with exercise coordinator SQNLDR Mark Barry (CO-SACTU), SACTU instructor SQNLDR Lou Desjardines, RASEC Liason Officer FLT LT Sam Hasenbosch[1], Gerry Bluett and Jack McCaffrey of Novonics Oceania and Brett Mobsby of Calytrix Technologies in the development of the evaluation.

1.2 Strategy and Design

The strategy adopted for this evaluation of the Mentor software was based on Kirkpatrick's (e.g. 1987) model of training system evaluation. Kirkpatrick's model of training system evaluation is a four-dimensional model. According to the model, a comprehensive evaluation of any training system should take into account the four factors of Reactions (of assessors and students to the training), Learning (what performance changes take place during training), Behaviour (transfer to on-the-job performance), and Results (in terms of the match between training outcomes and organisational goals). The first and second dimensions, namely student and assessor reactions, and performance change during the training event, were targeted for assessment here. The first dimension, student and assessor reactions to the Mentor system, was evaluated via qualitative analysis of transcripts and recordings generated during structured interviews with exercise participants. Information arising from this analysis is presented in Section 2. The second dimension, performance change, was evaluated via quantitative analysis of the Mentor performance ratings obtained during the exercise. Information arising from this analysis is presented in Section 3. When considered together, analyses along these dimensions speak to a broad spectrum of issues, from user acceptance and system usability to perceived strengths and weaknesses and the potential for the system to contribute to desired changes in student behaviour.

The design of the evaluation was developed in collaboration with exercise coordinator SQNLDR Barry. The planned exercise schedule consisted of 12 approximately-hour-long sessions in the SACTU air defence ground environment simulator (ADGESIM). These sessions were grouped into blocks of three, with each block taking place during either the morning (approx 0830-1130hrs) or afternoon (approx 1330-1630hrs). The exercise ran for two days. Each of the three simulator sessions in each block was manned by a different ABM team, though there was some crossover of personnel between teams. Of the three

[1] Sam Hasenbosch has since retired from the RAAF and taken up a position with DSTO Air Operations Division, Melbourne.

teams formed for the purpose of the exercise, one was defined as the Test Team (TT) and another as the Control Team (CT)[2]. The third team was observed, but was not included in the evaluation. Teams consisted of four operators, including three fighter engagement zone (FEZ) controllers and a WD. The WD acted as the team leader. While there was some sharing of roles between the FEZ controllers in each team across sessions, the WD maintained supervision of the team throughout the exercise. The schedule that was planned prior to the exercise is shown in Table 1 below (but see note below the table for changes to the actual schedule). Scheduling issues related to simulator and personnel availability meant that the TT and CT could not be assessed on all occasions that they were in the simulator. Instead, these teams were observed during the sessions indicated by grey-filled cells in Table 1. Both the TT and the CT were observed during their first and last sessions. In addition, the TT was observed on one occasion mid-exercise. The events included in the exercise depicted a scenario of gradually increasing hostilities. Therefore, on Day 1 and on the morning of Day 2, each hour-long session included different events. However, on the afternoon of Day 2 all three sessions were identical. This was to provide a fair comparison across teams at the conclusion of the exercise.

Table 1. Exercise Schedule

	Hour 1	Hour 2	Hour 3
Day 1, AM	Test Team	Control Team	Team 3
Day 1, PM	Control Team	Team 3	Test Team
Day 2, AM	Team 3	Test Team	Control Team
Day 2, PM	Test Team	Control Team	Team 3

Note: The CT and Team 3 sessions scheduled for Day 1 AM did not take place due to technical issues with the simulator. For the same reason, the TT session scheduled for the start of Day 1 actually took place around two hours after its planned start time. See Section 3 for a discussion of impact of this arrangement.

During their simulator sessions, the TT was assessed using the Mentor software and was then provided with feedback as a team via the Mentor tools in the form of handouts and AARs structured around stoplight reports. The TT took part in an AAR structured around Mentor stoplight reports at lunch time on both Day 1 and Day 2 of the exercise. At the end of both days they received feedback in the form of Mentor student handout reports. The CT was assessed using the Mentor software so as to provide comparison data. However, they were not provided with any Mentor feedback products. Two assessors took part in the evaluation. Due to scheduling and availability issues it was not possible to have both assessors assigned for all sessions and both teams (an arrangement which would have allowed an examination of the inter-rater reliability of the Mentor measures that were used). Instead, one assessor worked with the TT throughout and the other with the CT[3]. A tablet PC with the Mentor software installed was sent to the exercise coordinator approximately two months prior to the exercise to enable the assessors to familiarise

[2] Unfortunately, due to the availability of personnel, one member of the TT was also required to act as a member of the CT. While this was clearly undesirable from an experimental design point of view, it was unavoidable.

[3] This arrangement had a negative impact on the conclusions that could be drawn in regard to the performance differences between the CT and the TT. This issue is discussed further in the Sections 3 and 4.

themselves with the hardware and software. Also, a familiarisation and planning session was held the day before the exercise began.

During simulator sessions assessments were made against objectives and measures developed through collaboration between FLTLT Hasenbosch, SQNLDR Barry, SQNLDR Desjardines and the DSTO human factors team. As scenario events for this exercise were planned separately from objectives and measures, a relatively generic set of objectives and measures, which could be applied to a wide variety of scenario events, was generated. These were assembled into a means-ends hierarchy[4] with tactical-level Australian Joint Essential Tasks (ASJETS – Tactical Tasks; McCarthy, Kingston, Johns, Gori, Main & Kruzins, 2003) at the highest level and observable ABM team behaviours at the lowest level. During simulator sessions, assessors rated observed behaviours using a four-point scale that was based on typical SACTU performance-assessment practice. Scale points were associated with the verbal labels; SATISFACTORY (SAT), MARGINAL (MARG), UNSATISFACTORY (UNSAT), and UNRATED. The hierarchy of objectives and measures used in this evaluation can be found in Table A1 in Appendix A. In the sections that follow, the data arising from the CTT exercise are described and discussed. Data pertaining to student and assessor reactions to the Mentor tools are presented first, followed by data pertaining to changes in performance of the ABM teams over the course of the exercise.

2. Participant Reactions: Qualitative Analysis

Six structured interviews were conducted immediately after the conclusion of the last day of the exercise; one with each of the two assessors involved in the evaluation and one with each of the members of the TT. The aim of these interviews was to record the reactions of the assessors and students to the use of the Mentor tools for team training. For the assessors, the interviews contained questions designed to raise discussion in six areas, namely, (i) the data entry tool, (ii) handwriting recognition, (iii) format of the stoplight reports and handouts, (iv) objectives and measures, (v) the feedback provided to students during debrief, and (vi) teamwork concepts. For the students, the interviews contained questions designed to raise discussion on (i) the format of the stoplight reports and handouts, (ii) objectives and measures, (iii) the feedback provided to students during debrief, and (iv) teamwork concepts. The students were not asked about aspects of the software and hardware interface as they did not interact with the Mentor system directly.

A two-step process aimed at summarising views across participants was used to analyse responses to the structured interviews: First, two researchers independently listened to the interviews and recorded the themes that emerged from interviewee responses. A theme was defined as a common view, attitude, opinion, or judgment regarding an aspect of the

[4] Vicente (1999) describes means-ends hierarchies as those in which each node is an end that can be achieved by the nodes which link to it from below, and a means that can be used to achieve nodes to which it links above. As one ascends a means-ends hierarchy, the reason “why” each node exists is given. As one descends the hierarchy, nodes below reveal “how” each node is achieved.

way the Mentor system was used in this exercise. Themes were recorded if they were raised by more than one assessor or more than one student. Second, the researchers discussed the outcomes of their independent analyses and arrived at a consensus on a set of common themes. The themes to emerge during the interviews are presented in Tables 2 and 3. Table 2 summarises the themes raised by the assessors and Table 3 summarises the themes raised by the students. In each table, themes have been presented under the six aspects of the evaluation that were used to structure the interviews. As described above, the students had no direct experience with the first two categories and as such these columns in Table 3 have been omitted. Interviewees highlighted both areas of perceived strength and areas where the approach could be improved. These have been presented separately in the tables. Tables 2 and 3 also contain pointers to parts of Section 2 where the themes arising from the interviews are discussed in more detail.

Table 2. Themes raised during structured interviews with assessors

	Tool: **Navigation and Rating**	**Tool:** **Handwriting and Comments**	**Tool:** **Stoplight Reports and Handouts**	**Content:** **Objectives and Measures**	**Content:** **Performance Feedback**	**Content:** **Team Approach**
Perceived Strengths	- Easy to use, navigate (see theme 1) - Link between serial events and objectives provides prompt to assessor (see theme 2)	- Handwriting recognition generally good (see theme 5)	- Drill down functionality good for debrief (see theme 9)	- Objectives & measures generally good (see theme 12)	- Rating scale easy to understand, conforms to standard approach (see theme 15)	- Team approach is important (see theme 22) - Team dimensions easy to understand (see theme 23) - Team-level assessment and feedback under-emphasised (see theme 22)
Observations and Suggested Improvements	- Weight of tablet PC too great to carry for long periods of time (see theme 3) - Screen real estate can be an issue (see theme 4)	- Not obvious when in handwriting mode (see theme 6) - More eyes-down and effort required than paper & pencil (see theme 7) - Need the ability to draw diagrams (see theme 8) - Screen real estate issues (see theme 4) - Need dictionary of air defence terms (see theme 5)	- Need to display comments against all levels of objectives in reports & handouts (see theme 10)	- Smaller number of more tailored objectives and measures required (see theme 12) - Definition of serials could be better: possible mix between event categories and temporal sequence (see theme 13) - Run time addition and removal of objectives and measures desirable (see theme 14)	- Weightings should be applied to emphasise important objectives (e.g. safety, tactical) (see theme 16) - The tool should facilitate comparisons across sessions (see theme 17)	- Team approach suited to *learning* rather than *assessment* (see theme 24) - Teamwork dimensions could be more tailored to air defence context (see theme 12)

Table 3. Themes raised during structured interviews with students

	Tool: Stoplight Reports and Handouts	**Content: Objectives and Measures**	**Content: Performance Feedback**	**Content: Team Approach**
Perceived Strengths	- Hierarchical objective structure clear and easy to understand (see theme 9)	- Objectives & measures generally good (see theme 12)	- Rating scale easy to understand, conforms to standard approach (see theme 15) - Timeliness of feedback is a key advantage (see theme 18) - Useful for future training courses/ exercises (see theme 19)	- Team approach is important (see theme 22) - Team dimensions generally easy to understand (see theme 23)
Observations and Suggested Improvements	- Need to display comments accurately and against all levels of objectives (see themes 5 & 10) - Large number of displayed objective levels or unrated measures can be distracting (see theme 11)	- Smaller number of more tailored objectives and measures required (see theme 12)	- Students must understand the tool and the process to achieve maximum benefit (see theme 20) - Scores show what went wrong, comments show how to fix it (see theme 21)	- Team approach should be an adjunct to individual assessment and feedback (see theme 25)

It was clear from responses to the structured interviews that assessors and students saw considerable value in both the Mentor tools and the team training approach embodied in them for the purpose of this exercise. Mentor was seen as an easy way of providing structure, objectivity of assessment and timely feedback to students, while teamwork and team skills were seen as important aspects of performance that are currently underemphasised. These and other points raised by assessors and students during the interviews highlighted concepts which require further discussion. To this end, each of the themes summarised in Tables 2 and 3 are considered in more detail below. Recommendations are presented for each point in order to indicate where further investigation or development of the approach should be focused.

2.1 Discussion of Themes from Structured Interviews

1. Tablet hardware and Mentor software is generally easy to use and navigate
The assessors were generally satisfied with the Tablet PC and the Mentor software interface. They found the tablet and pen easy to use and had little difficulty rating performance and navigating between serials. Although they were largely satisfied, some issues were raised relating to the pen. One assessor reported that the right-click button on the pen, which was positioned on the pen's shaft, was badly placed and could be pressed accidentally. When this occurred, ratings could not be made and the writing tool could not be selected. Also, the spare pen was found to be too small to be used comfortably.

Recommendation: While initially frustrating, problems with pressing the right-click button on the pen are likely to decline as familiarity with the pen increases. However, if this issue is found to recur, it may be necessary to acquire a pen on which the position of the button is less problematic. The right-click button is not frequently utilised in the context of the Mentor software and there is therefore no real requirement for it to be readily accessible.

2. The links between serial events, objectives, and measures provides prompts for assessors
It was noted that the presence of the measures on the DET that were tailored to serial events prompted assessors to rate specific aspects of performance for each different serial. By design, the Mentor tool allows specific objectives and measures to be attached to particular serials. This helps assessors to stay focused on relevant aspects of performance in relation to specific events, rather than generic aspects of behaviour. This is useful as it ensures that student assessment is targeted and allows assessors and students to develop an understanding of the student's performance profile across a range of tasks. In addition, prompting assessors to rate students on specific measures increases the objectivity of performance, as ratings and comments may be less likely to be influenced by global impressions (e.g. halo effect).

While a significant amount of effort was made to tailor objectives and measures to serial events during the CTT exercise, these elements of the training event were not as well matched as would ideally be the case. This was evident in the generic nature of some of the measures which came about due to the method by which the Mentor tool was populated. The scenarios were created first, and were then segregated into serials. The objectives and measures were created in parallel and relevant measures were then

attached to serials. Optimal use of the Mentor tool would involve the scenarios, serials, objectives and measures being created concurrently. This should result in an association between serials and measures that is tighter and more focused on the specific objectives and behaviours of the team undergoing assessment.

Recommendation: The utility of the Mentor tools will be maximised if the scenarios, serials, objectives and measures are created concurrently, as this is likely to increase the specificity of the measures obtained and the feedback provided.

3. The weight of the tablet PC is too great to carry for long periods of time

The manufacturer's advertised weight for the Tablet PC used in this exercise is 1.75kg. While this is relatively light, the assessors reported that the Tablet PC was too heavy to carry for prolonged periods of time. The effect of the PC's weight was different for the two assessors. One assessor found it awkward to carry the PC around at all, and so opted to position it on a table and to rate student performance from a seated position. This assessor observed the team from a remote position while viewing activity on a tactical situation display and listening to communications made on the radio channels. The other assessor found that the PC afforded somewhat greater mobility. However, it was still found to be awkward to carry for extended periods. This assessor worked for the most part with the tablet in their lap or cradled in one arm and preferred to observe from a position near the ABM team where visibility of the team's behaviour and interactions, and of the communication between team members, was greater. To the extent that the PC hardware led to these differences in assessment style, this represents a problem for standardisation of assessment.

One potential solution to this problem would be to use a personal digital assistant (PDA) rather than a Tablet PC for presenting the DET (e.g. Clark, Lenne, Robbie, Ross, Ryan, & Zalcman, 2003). However, this approach would come at a cost in the form of a dramatic reduction in available screen space (around 2.5 times less space). The issue of screen space was also highlighted by the assessors during the structured interviews and is discussed below.

Recommendation: Ideally, the weight of the device used for presenting the DET would not constrain assessor rating behaviour at all. However, a trade off must be struck between the weight of the hardware and available screen size. The value of screen space was repeatedly emphasised throughout the interviews, and therefore downsizing to a PDA does not, at present, appear to be a viable option. Given that the weight of Tablet PCs is likely to reduce over time, this may become less of an issue in the future.

4. Screen space in the DET interface is at a premium and must be managed carefully

It was clear from the assessors' responses that DET screen space should be managed carefully when designing the tool's interface. Given a device of fixed screen size it is clearly necessary to economise on DET screen space. However, the requirements of the users should be taken into account when making decisions on what to display and how to display it. An example of the current DET economising on the use of screen space in a way that users judged undesirable is the way large numbers of measures and lengthy comments are displayed. Currently, lengthy comments and measure names are

abbreviated such that only the first and last portions of the text are displayed in the main DET window. One assessor felt that it was important for entire comments to be readable, as a prompt to memory, after the text-entry box has been minimised. This assessor also felt that all measures relevant to a given serial should be displayed on a single screen, eliminating the requirement to scroll. However, clearly this demand must be traded off against other demands such as those relating to the number of available measures and text size. A satisfactory balance between the competing desires to display a great deal of information and to fit it all onto one screen may be difficult to strike. However, if objectives and measures are more closely tailored to the scenario events than was the case in this exercise, it may be possible to reduce their number, thereby reducing demands on screen space.

Recommendation: The suggestions that comment boxes expand to display the entire comment contained in them and that all serial measures be contained in a single screen could be useful to explore as ways to enhance the DET interface. In order to strike a balance between these competing demands, an upper limit on comment expansion could be set based on the rule that comments be as large as possible while permitting all measures to be displayed on a single screen. Whatever strategies are adopted in the interests of making most effective use of DET screen space, they should be based on a solid understanding of user requirements.

5. Handwriting recognition was found to be generally good, but could be improved

Both assessors gave positive evaluations overall of the accuracy of the handwriting recognition software used to record comments (Microsoft Tablet PC Input Panel version 1.7). They found it to be surprisingly accurate, even when the quality of handwriting was poor. Although some errors of recognition did occur, the intent of the comments was usually apparent.

In terms of workload, both assessors found that handwriting notes on the PC required more effort and concentration than writing with pen and paper. They reported needing to concentrate more on the quality of their handwriting and to monitor whether it was being translated accurately. In particular, both assessors also found it difficult to modify or delete words that had been incorrectly recognised. One assessor noted that lack of familiarity with the tool, difficulties with using the handwriting function and the requirement to rate performance on a large number of measures caused a reduction in the frequency and depth of comments made during the exercise. As assessor comments are important for student learning, factors reducing the frequency, depth or quality of comments are likely to negatively impact training outcomes. Fortunately, it was reported that this impact was at least partially ameliorated by familiarity with the tool.

Nevertheless, there did appear to be some consistent problems with the handwriting recognition. One of the main problems related to the context-sensitive nature of the word and sentence recognition. One assessor reported that the translation of a word would change depending on the words surrounding it – sometimes going from correct to incorrect. A related problem was that letters were recognised in the context of other letters in the same word. It was reported that if the software interpreted the first letter of a word incorrectly, the entire word was almost guaranteed to be translated incorrectly. One

assessor found that most problems of this type occurred when words began with the letters 'R' or 'C'. The context-sensitive recognition feature may be useful in other environments where whole words, common phrases and grammatically correct sentences are the norm. However, in the air defence environment assessors often record comments in a format that is grammatically incorrect, using abbreviations and sentence fragments. This was found to reduce the accuracy of handwriting recognition. A related point is the participants' suggestion that a dictionary of air defence specific terms and abbreviations should be incorporated into the handwriting recognition tool. As in most work environments, there are a large number of acronyms and specialist terms used by air defence personnel that are unique to this environment and thus do not appear in a general dictionary. The students and assessors felt that such a dictionary would improve the accuracy of the handwriting recognition software.

There are clearly benefits of being able to provide students with feedback immediately following assessment that conveys their performance on a range of measures and suggests methods of improvement. These benefits will be discussed later. In its current form, the handwriting tool seems to be capable of conveying the comments made by assessors in a form that is interpretable, albeit not always entirely accurate.

Recommendation: Familiarity with data input via the DET is likely to alleviate some of the problems discussed in this section. The context-sensitive word and sentence construction logic appears to reduce the accuracy of handwriting recognition when the dictionary does not contain specialist terms. Therefore, word recognition may improve if a dictionary of air defence specific terms, acronyms and abbreviations is included. A training feature, in which the handwriting recognition tool is trained to recognise an individual's writing style as well as particular terms, would likely be advantageous. In the absence of such a feature, assessors could be directed to modify their writing style to form problem letters and words in a specified way. However, this would increase workload unless assessors were highly practiced. Alternatively, they could use the letter-by-letter word recognition feature. This has the advantage of being more accurate, but is likely to reduce the speed with which comments can be recorded. Another option would be to record the handwriting for later presentation in bitmap form, without converting to text. This option may be particularly useful during high activity phases when the assessor may not have the luxury of the 'eyes down' time to monitor the accuracy of handwriting recognition.

6. It is not obvious when the DET is in handwriting mode

Both assessors reported that they were sometimes unsure whether the DET was in handwriting recognition mode and whether comments were being inserted at the correct point. They reported that there was no obvious feedback to indicate the mode or the input position and they found that as a result, comments were not always attached to the correct measure. For this reason, one assessor reported that it was easier to record comments on the overall serial notes page during the scenario and then edit and insert these comments under the appropriate measures at the scenario's conclusion. While this is a straightforward workaround, the method may be problematic in that it unnecessarily increases reliance on the assessors' memory for events which took place during the exercise. Reliance on memory for events can be risky as memory has been shown to be highly susceptible to influence, error and bias (e.g. Wells & Loftus, 2003).

Recommendation: This issue is likely to become less problematic as assessors become more familiar with the tool. However, it would be a simple matter to provide additional feedback in the DET interface to reduce the risk of mode confusion. Such feedback should serve to highlight the measure to which comments are being attached as well as making very clear the active/passive status of the text entry box.

7. Using the Tablet PC and Mentor DET required more cognitive effort and 'eyes down' time than paper and pencil

The assessors commented that use of the handwriting tool required more time to be spent looking down at the PC than would be the case if they were writing using paper and pen. Assessors reported the need to keep looking down to ensure that they were writing in the right location, to ensure that their handwriting was being correctly recognised, and to make corrections when failures of recognition occurred. The extra time spent looking at the PC and the extra cognitive effort involved in entering data could have been better spent observing activity and monitoring and interpreting team interactions.

Recommendation: Much of the effort involved in using the Tablet PC/DET combination arose from the use of handwriting recognition. A suggestion was made earlier (see point 5 above) regarding capturing handwriting in its raw form as a bitmap, rather than converting to text. Converting assessors handwriting to text has potential benefits regarding data analysis, for example the ability to search databases of converted comments for particular keywords. However, it is not clear whether such functions will actually be built into future systems, or whether the users of such systems will find them beneficial. The option to capture handwriting as a bitmap rather than converting to text, and other options which could reduce assessor workload, should be explored.

8. The ability to draw diagrams and make them available to students is highly desirable

The events which take place in air defence scenarios have a strong geometric character, involving interactions between entities that take place in a volume of space and time. Explanations of these events and suggestions for action that rely heavily on spatial relationships are likely to be easier for students to understand when supplemented by a graphical representation. For this reason, a drawing function would seem to be a very useful addition to the Mentor tool. Both assessors and one of the student participants commented that it would be useful to incorporate a drawing function into the Mentor DET. Assessors could access the function during the assessment period and use it to draw diagrams that illustrate their comments and suggestions for improvement. These diagrams would be exported to the feedback products along with ratings and comments and could be displayed during debrief to promote students' understanding of where they went wrong and how to improve their performance in the future.

While a drawing function would provide for a more direct representation of the geometric relationships inherent in air defence contexts than would spoken or written language, codification of scenarios into two-dimensional diagrams would still involve a cognitive transformation. The third dimension of space and, in particular, time would not be represented. An even more direct representation of the scenario could be made available through the use of AAR playback tools. AAR tools are available which allow playback of

events recorded from simulator sessions. Scenarios played back through these tools can typically be explored by zooming and rotating, and the temporal aspects of the scenario can be preserved, or indeed manipulated by pausing, rewinding, and playing in slow motion to enhance understanding.

Recommendation: A drawing tool, at the very least, would dramatically improve the utility of the Mentor tool for the air defence context. In addition to the drawing tool it would be very useful either to include a scenario recording and playback feature, or for users of the Mentor software to supplement their AAR through the use of other applications that provide such functionality. In point 4 above, the issue of screen space was discussed. If a drawing function or similar is implemented, it would not be advisable for drawings to be displayed permanently on the DET screen or in the feedback products as a default as they would occupy too much space. A windowing solution is likely to represent the best option.

9. Drill-down functionality and hierarchical structure of Stoplight reports was useful

The assessors and most of the students found the hierarchical structure and drill-down functionality of the stoplight reports to be extremely useful. Some of the impact was lost when students were first presented with the stoplight report because the structure and content of the reports was not explained to them in detail prior to the AAR and the session was very rushed. This created some confusion for students as to how the information in the stoplight reports was organised and what the scores represented. After being properly briefed on the stoplight reports most students found the method of presentation to be useful. Also, most students reported that the scores and colour-coding of stoplights made it easy to see what was done well, what was done poorly and which aspects of team performance required improvement.

Recommendation: The stoplight reports should always be properly explained before being presented to trainees. When it was explained, the stoplight report was evaluated as very useful. In the versions of the reports used for this exercise, four levels of abstraction were hard-coded into the reports. However, the objectives and measures used to structure performance assessment included only three levels (see Table A1). This meant that one level had to be repeated in the stoplight reports, making the structure seem more complicated than it needed to be. The inclusion of this extra level increased the potential for confusion in the students. While this problem could have been remedied by having new report templates generated, this is currently not something that can be easily done by the end user. The report formats should be made more flexible, such that the number of levels best suited to the context at hand can be specified by the end user.

10. The DET and reports should allow comments to be made and displayed against all levels of events and objectives

The Mentor DET currently has the facility for assessors to record comments against sessions, serials, roles, and measures. This is an important function as it allows assessors to include amplifying information for student feedback and it can also serve to fill gaps where training design has not identified objectives and measures for all relevant behaviours that are observed. There is, however, no facility to record comments at the levels of objective categories or objectives. The assessors reported a desire to record

comments at all of these levels. It was also noted that comments that were recorded in the DET were not always available or easy to find in the student handouts and stoplight reports. In the handouts, only comments recorded against the measures are included. In the stoplight reports, it is easy to find comments made next to each measure, as they are automatically displayed when users drill down to view the ratings made against each measure. It also easy to find the comments made against each serial, as this can be done by clicking on the serial's hyperlink. It is not, however, easy to find where the overall session comments are displayed. The word 'Overall' appears in the top left corner of the stoplight reports, but does not feature a hyperlink to comments. There are many underlined headings in the stoplight reports, only some of which are working hyperlinks. In addition, there does not appear to be an easy way to discriminate hyperlinks that contain comments from hyperlinks that do not. After completing a session, an assessor may not remember where comments have been recorded. A visual prompt at the interface to remind assessors of where they have recorded comments would be very useful.

Recommendation: The facility for assessors to make comments against sessions, serials, teams, objectives, sub-objectives and measures should be included. All comments made should be easily accessible in the handouts and stoplight reports. Only working hyperlinks should be underlined in the stoplight reports to avoid confusion. Lastly, hyperlinks should only exist if a comment has been recorded. For example, if overall comments have not been recorded for Serial 1, the Serial 1 label should not be a hyperlink. This will ensure that assessors will not open a number of empty hyperlinks in the search for an elusive comment.

11. Displaying a large number of unrated and uncommented objectives and measures in reports can be distracting

The Mentor system facilitates the provision of feedback to students in the form of (i) stoplight reports, and (ii) student handouts. These two feedback products contain the same information; however the display format of each is tailored to its intended use. The stoplight report is intended for use as an after action review tool, while the handout is intended for use as part of a take-home package to encourage students to reflect on their performance and that of their team (see Figure 2 for an example of each). It is possible for measures to be unrated within the Mentor data entry tool. This typically happens when assessors see no behaviour relevant to the item in question during the exercise being assessed. Currently, the Mentor feedback products display all measures – both rated and unrated. Feedback from students taking part in this exercise indicated that the inclusion of objectives that were unrated and had no comment against them in feedback products was distracting. While this was an issue for both feedback products, it was less so for the stoplight report, as this was used in conjunction with an assessor-led discussion which served to guide the students' attention.

Recommendation: An option should be provided within the Mentor tools to export only rated or commented objectives to the feedback products if it serves assessment and feedback purposes to do so. This will assist in directing assessor and student attention to those aspects of performance that were actually observed. It may be valuable to explore the utility of displaying a value alongside aggregated stoplights in the stoplight report which indicates the proportion of measures underneath that stoplight which have actually

been completed. This would provide information about how many of the available measures actually fed into the aggregated result at higher levels of the hierarchy. This information could be relevant in determining the way assessors interpret aggregated results.

12. The objectives and measures defined for this exercise were generally appropriate, but required refinement

The students and assessors found the objectives and measures defined for this exercise to be generally appropriate, but commented that refinement would be required if the tool were to be adopted. They found the set of measures to be too general and not tailored specifically to the missions being run. The assessment was seen as somewhat superficial and not as beneficial as it may have otherwise been in terms of learning. It was suggested that greater analysis of the team interactions would be required if teamwork training was to be effective. Those interviewed agreed that there were too many measures. It was felt that the quality of assessment would benefit from the inclusion of a smaller number of measures that were perhaps slightly broader, but covered issues that were more relevant to the particular scenario and to the air defence context. It was commented that if Mentor was to be used for training in an operational context, it would be extremely important for the right measures to be included as behaviours that were not included as measures would tend not to be discussed in the debrief. The objectives and measures are therefore among the most important and influential aspects of the tool and their definition and development should be considered one of the key inputs required to maximise the effectiveness of the system.

Recommendation: Considerable effort will be required to define and maintain the Mentor objectives and measures if the tool is to be used on an ongoing basis. Organisations seeking to use Mentor to support training events should pay close attention to the question of how this material is to be defined and managed, as this is likely to represent both a major investment and a major determinant of the quality of the outcomes that are achieved. The process of defining and managing objectives and measures is likely to be time consuming and expensive, and it should not be considered a once-off undertaking. For objectives and measures to remain relevant and useful, they should be reviewed and revised on a regular basis in the light of operational priorities and lessons learned. The definition and ongoing refinement of objectives requires input from training experts and subject matter experts with adequate experience and expertise, as well as an appreciation of the specific training outcomes under consideration.

13. Tailoring Serials to Context

It was the view of both assessors that the serial structure used during the simulation exercise could be refined to be more suitable to the air defence context. The serials defined in Mentor for this exercise were based on clusters of time-sequenced events, with the serial start points coinciding with the appearance of an entity or the onset of some system or environmental state. This has been the manner in which the tool has been used previously, and with some success, in the maritime domain. However, the rapidity with which the situation can develop in the air domain meant that during the exercise, serial events often merged into one another. When this occurred, assessors were required to shift their attention between events and navigate the DET between serials.

A clear example of this happening during the simulation exercise involved enemy aircraft undertaking air-to-air refuelling. If Serial 1 is defined as the appearance, transit to tanker, and later transit to target of a group of hostile aircraft, and Serial 2 is defined as the appearance and transit of another group directly to their target, the ABM team may be required to respond to Serial 2 around the same time as, or even before, Serial 1. This is because the process of tanking before proceeding to target can cause the first group to take longer than the second to reach an area being controlled by the team. When events related to two different serials occur roughly simultaneously, behavioural observations related to the two serials can be confounded. This can negatively affect the specificity of the information available for trainee assessment and feedback. When serials occur out of order with the DET structure, assessors are required to navigate the DET pages back and forth to find the appropriate objectives and measures for each serial. This can increase assessor workload and may have a negative impact on the quality of observations, ratings, and comments.

Recommendation: Alternative methods of defining serials for the purpose of structuring training sessions should be explored in order to more closely tailor the Mentor DET and reports to the rapidly-evolving nature of the air domain. In order to maximise the validity of observations and minimise assessor workload, it may be necessary to explore the feasibility of defining serials in terms of the estimated time-sequence in which the team will be required to act rather than with respect to when an entity appears or the onset of a system or environmental state.

14. Run-Time Addition of Objectives and Measures

The DET used by the assessors for assigning ratings and comments during the simulation exercise (see Figure 1 for an example of the DET interface) was populated with objectives, measures, and serials that were defined in the lead-up to the exercise through consultation between the RAAF exercise coordinator, other ABM subject matter experts and DSTO human factors researchers. These elements of the exercise structure were designed to tap into important ABM taskwork and teamwork competencies, and to trace back, via a means-ends hierarchy, to high-level organisational goals of the ADF (i.e. the Australian Joint Essential Tasks; ASJETS). Among the major long-term benefits of using a system such as Mentor for objectives management in this fashion is that it provides a framework for defining objectives and recording performance against them across many exercises. Over time, this would provide data which speak to the development of operational readiness and provide an 'audit trail' that facilitates the identification of organisational strengths and weaknesses. This information could be used not only to ascertain with a degree of rigour the current state of the organisation relative to goals, but also to tailor training events to precisely target perceived shortcomings.

The assessors involved in the simulation exercise reacted positively to the objectives and measures that were defined for the exercise in the Mentor tool. However, they both felt that the ability to add and remove objectives and measures in an ad-hoc fashion *during exercise run time* would be a desirable enhancement to the software. The rationale behind this desire was that it was viewed as almost impossible to anticipate all of the events and behaviours that may arise during an exercise of this kind. The ability to modify the objectives and measures available to the assessors for a given exercise, or serial within an

exercise would provide the flexibility to assign ratings and comments against behaviours that were observed, but which had not been anticipated during preparation for the exercise.

The current Mentor software would in fact allow assessors to update the DET with objectives from the main Mentor tool during an exercise. However, this would involve the generation of a new DET form, which is a relatively cumbersome operation involving the transfer of data between two of the software tools. Given the rapid pace with which scenarios can develop in the air defence context, assessors would have to sacrifice a significant amount of observation time in order to complete that operation. If run-time addition and subtraction of objectives and measures was deemed to be a valuable enhancement to the tool, a new method (e.g. some action akin to dragging and dropping via a graphical user interface) would need to be developed to simplify the process.

However, regardless of implementation, run-time addition and subtraction of objectives and measures could have negative impacts on the quality and utility of the assessments made using the Mentor tool. The addition and removal of different objectives and measures at different times is likely to complicate the comparison of performance across similar teams and exercises. This would detract from the ability of the system to facilitate the assessment of the development of operational readiness and the identification of organisational strengths and weaknesses. In simple terms, one may end up trying to compare apples and oranges. Another potential problem with run-time addition and subtraction of objectives and measures relates to the pervasive psychological bias known as confirmation bias (e.g. Wickens & Hollands, 2000). This term refers to the tendency for all humans to seek out and attend to information which confirms initial impressions, while ignoring or otherwise downplaying information which is contrary to initial impressions. With run-time addition and subtraction of objectives and measures, assessors could form a global subjective impression early in an exercise, then proceed to add objectives and measures which provide evidence to confirm that impression, while removing objectives and measures which provide evidence against it. The pervasive nature of cognitive biases such as this suggests that assessor experience and expertise may provide little protection from such outcomes. Because of the effect that confirmation bias could have on assessments, if it was used in this way, the Mentor tool could lead to greater, rather than less, subjectivity in assessment.

Recommendation: If run-time, ad-hoc addition and subtraction of objectives and measures is to be supported in the Mentor tool, two issues should be addressed. First, the implementation of this functionality will need to be streamlined to reduce the amount of 'eyes-down' time required by the assessor. Second, users should be wary of applying this functionality and its application should be controlled in a stringent fashion to preserve the quality and utility of the information generated via the system. The potential benefit of flexibility provided by this functionality is in all likelihood outweighed by potential costs to structured objectives management, readiness evaluation, and the validity of the observations made using the tool. As a method for recording unexpected behaviours, the comments facility provided by Mentor appears far less problematic.

15. Rating scale was easy to understand and conformed to standard approach
Within Mentor a rating scale must be assigned to each measure to facilitate performance assessment. During the exercise, assessors assign ratings against measures using the defined rating scale using the DET interface. The number of points on the rating scale, the values assigned to each point and the verbal labels used to describe each point can be configured by the user. For the purposes of the exercise reported here, a set of four verbal labels familiar to the SACTU assessors was adopted for the Mentor DET (see Section 1.2 for a description). These verbal labels were easily instantiated in the software and were evaluated as familiar and easy to apply in this context by the assessors.

However, this approach could be refined in future applications of the Mentor tool. Verbal labels of the kind described above are very abstract and as such are open to different interpretations by different assessors. This can reduce the reliability of assessments, as it is up to each individual assessor to arrive at a judgement about exactly what each label means and what constitutes behaviour worthy of each descriptor. The most commonly-applied solution to this problem is to provide a set of behavioural anchors for each scale point, or at least a subset of scale points (e.g. the uppermost and lowermost points). Scales which contain such anchors are known as behaviourally-anchored rating scales (BARS). By assigning a description of typical behaviours to points on the scale, assessors can develop more consistent expectations of how behaviours should be rated. A good example of a BARS approach in the military training domain are the generic measures of performance for distributed mission training developed by the Canadian defence research organisation DRDC (Matthews & Lamoureaux, 2003). These measures contain both behavioural anchors for each scale point and a list of behaviours relevant to each item to guide the assessors observation. An example of one such measure is provided in Figure 3 below.

Recommendation: The reliability and validity of assessments provided via the Mentor DET could be enhanced by including support for the provision of behavioural anchors against scale points. Given the sheer number of anchors that would need to be developed, the inclusion of these would represent a non-trivial addition to the effort required to prepare a Mentor database in support of a given training exercise. This would also have an impact on screen space, as more area would be required to display such items. However, with some development of the software interface the latter problem could be alleviated. Also with reuse of material from previous exercises, the effort required to develop measures and associated behavioural anchors will diminish over time. The payoff for undertaking this development effort is likely to be high as formatting measures in this way can be expected to enhance the quality of recorded observations.

1. Poor	2. Marginal	3. Standard	4. Very Good	5.Exceptional
13. MISSION AND GOAL AWARENESS: RE-ESTABLISHES MISSION GOALS, DETECTS AND RESPONDS TO CHANGES IN MISSION PICTURE				
Fails to detect changes in mission situation following engagement Poor resumption of nav plan/join of formation following engagement	Recognises change in mission picture Unsure of appropriate COA in response to change Unsure of mission goals at various points in mission	Recognises changes in mission situation and adjusts appropriately	Integrates information to quickly recognise changes in mission picture Rapidly updates plan and communicates changed picture and plan	Anticipates changes in mission picture Has contingency mission goals
Look for: Maintaining a broad scan of info sources (i.e. radio, hud, outside, etc.). Ability to comprehend current changes in the tactical environment. Ability to anticipate the effect of current changes in the tactical environment on future events in the mission. Ability to develop contingency COAs in response to the potential impact of current mission events Appropriate assessment of tactical situation and use of defensive, offensive and neutral manoeuvring tactics **Observations:**				

Figure 3. An example of a behaviourally anchored rating scale (BARS) item from the DRDC list of generic measures for distributed mission training (DMT)

16. Weights should be applied to emphasise important objectives and measures

The Mentor tool provides for the assignment of weights to measures and objectives, which influence the impact that each has on the aggregated scores to which it contributes. When scores are aggregated within Mentor, those measures and objectives which have been assigned large weights will have a greater impact on overall scores than those measures and objectives which have been assigned small weights. This approach was generally supported by the expert assessors involved in the exercise reported here. In particular, they expressed a requirement to weight safety-critical items more heavily than other items in the assessment of overall performance. In the context of air defence team training, safety critical items include those related to factors such as aircraft separation standards, other aspects of airspace management, and the application of emergency procedures.

For this reason, the ability to emphasise some factors over others by assigning weights would appear to be a desirable function in Mentor. However, a problem with this function is that there currently exists no firm, empirically-validated basis upon which weights can be assigned. A simple approach would be to have the weights act in a binary fashion, with one weight applying to standard items, and another, heavier weight applying to mission or safety critical items. Indeed, this was the approach taken for the purpose of preparing after action review products during the exercise reported here. Safety critical items identified by the SACTU assessors were assigned a weight that was double that of the standard items. While this unsophisticated approach served well to demonstrate the weighting functionality to the assessors involved in this evaluation, it would not suffice as a long-term strategy. An alternative strategy would be to treat weights as a representation of each item's 'importance'. This would require that the importance of measures and objectives be established via empirical investigation involving a large number of expert assessors. In order to achieve this, one would have to first develop objectives and

measures, then survey a number of expert assessors within the field of study, asking them to assign an importance score to each item. The weight assigned to each item could be derived from the importance scores assigned by the surveyed assessors.

While this strategy would go some way towards managing the aggregation of scores from items that are considered to be more or less important than one another, it does not provide for maximum flexibility in the modelling of mission or safety criticality. In particular, the use of weights in this fashion does not directly model the conditional manner in which mission or safety criticality is sometimes conceptualised. For example, assessors may want to set up a situation in which an UNSAT rating on a particular safety-critical item or set of items will lead to UNSAT ratings propagating upwards to a certain level of aggregation in the overall assessment, regardless of other scores. They may or may not want the effect to propagate all the way up, so that the entire exercise is rated UNSAT. This is sometimes referred to as a critical failure. While this effect can be approximated using scores in the current Mentor implementation, a more direct and flexible way to achieve this would be to enable assessors to assign rules which override the aggregation of scores. For example, to accommodate the situation described above, a simple set of rules could be applied to ensure that any linked objective up to n levels of aggregation above critical measure x is evaluated to UNSAT if critical measure x is itself rated as UNSAT.

Recommendation: Consideration must be given during the preparation for any training exercise in which Mentor is to be used to: (i) whether weights are to be assigned to measures and objectives, and (ii) if so, how those weights are to be determined. A possible strategy would be to treat the weights as a measure of importance. This would involve having a number of expert assessors rate the importance of each item after they have been developed and are ready to be entered into the Mentor database. However, this approach could prove time consuming. A valuable enhancement to the current Mentor tool for the purpose of managing mission or safety critical objectives and measures would be the ability to define rule-based strategies for score aggregation which override the scoring system. While the effects of such rules can be approximated using the extant scoring system, a direct approach may provide greater utility.

17. The Mentor tools should facilitate between-session performance comparisons

Mentor currently provides for performance feedback to trainees via two output products; (i) the stoplight report, used primarily for after-action review, and (ii) the feedback handout, designed to encourage trainees to reflect on their performance outside of the exercise. Examples of the designs of these Mentor feedback products from this exercise are displayed in Figure 2. As can be seen from this figure, both feedback products contain information about objectives, measures, and associated performance ratings and comments, typically for a single training session. While these products were regarded by participants as generally useful, the potential for improvement in terms of tracking performance change over time was noted during structured interviews with assessors and students. In many situations, the provision of information regarding performance change could be valuable in order to provide positive reinforcement and to direct trainee's attention to aspects of their performance which require attention. While this could be achieved using the filtering functions in the current implementation of Mentor, this information was not provided in the exercise reported here. However, such information

may be quite useful in exercises where sessions of similar difficulty or complexity are repeated over time. An example of how such information could be displayed in the Mentor stoplight report would be to replace the coloured circles associated with each objective and measure (see Figure 2) with upward and downward pointing arrows for measures on which performance has increased and decreased respectively since the last similar training session. Alternatively, concurrent sessions could simply be displayed alongside one another to facilitate comparison.

Recommendation: Given that the raison d'etre of the Mentor suite of tools is to facilitate learning, and learning can only be assessed by examining changes in behaviour over time, evaluations of the worth of the Mentor tools should include consideration of the extent to which the software facilitates such comparisons. Consideration should be given to the utility of displaying performance change across sessions in a manner that is sufficiently clear to provide useful information for assessors and constructive feedback to students.

18. The timeliness with which feedback can be provided using Mentor is a key advantage of the system

The provision of timely, accurate, and relevant feedback is important for learning. One of the key benefits of the Mentor system is that performance feedback can be available in the form of the stoplight reports and student handouts within minutes of the end of an exercise. During the exercise reported here, the process of generating these feedback products after each simulator session generally took less time than that required to assemble all of the relevant personnel into the classroom for debrief. As a result, the products could be used to guide discussion of events and implications for future performance while students' and assessors' memories of the session were still very fresh. The students found this to be a particularly attractive aspect of the software. Whichever other aspects of the Mentor tools may prove to be worthwhile, the provision of such timely feedback is likely to be a key advantage.

Recommendation: One of the most promising aspects of the Mentor suite of tools is the ability of the software to facilitate the provision of rapid, detailed feedback to students. Users of the software should be aware of this strength and take full advantage of the package's after action review and take-home handout products to enhance learning.

19. The Mentor tool is likely to be useful for future training events

After seeing the Mentor tools for the two days of the exercise, students participating in the experimental team expressed the view that the software could be useful for supporting future training events of this kind.

Recommendation: Evaluations of the Mentor suite of tools should continue so as to identify both benefits of the system and areas where it could be improved or tailored more closely to the requirements of the RAAF. While not without its shortcomings, the evaluation effort described in this report could serve as a starting point for refinement of the system and as a model for future studies.

20. Maximum benefit can only be achieved if students understand the tools and processes involved in team-level assessment

Due to time constraints, there was very little opportunity at the beginning of the exercise to introduce the students to the Mentor software tools or the team assessment and feedback approach that was embodied within it. Explanations of the tools and the objectives and measures, including teamwork dimensions, were embedded briefly by the exercise coordinator in the general introductory session (attended by all participants) and by the test team assessor within after-action reviews as objectives and measures were discussed. This meant that the students in the TT were required to develop their understanding of the new material (e.g. teamwork dimensions, format of feedback products) as the exercise took place.

During structured interviews at the end of the exercise, some students indicated that they felt that their understanding of the material was less than satisfactory, and asked their interviewer to explain the Mentor tools, the teamwork dimensions, objectives and measures in more detail. While all expressed the view that the approach and content were straightforward when explained, they also indicated that their lack of clear understanding during the exercise could have hampered their learning. This is an important point for the planning of future activities of this kind. When students are not provided with a clear explanation of the structure and content of assessment and feedback, they will be required to devote significant cognitive resources to simply working out these issues in an attempt to interpret the information being presented to them. This may leave little in the way of cognitive capacity to be devoted to the more important issues of reflecting on performance, contributing to team discussions, and planning behaviour modifications for future sessions. What's more, without a good understanding of assessment and feedback processes, students may suffer from low motivation and become disengaged from the learning experience. This could be a particular problem for team training of the kind examined here if student attitudes towards the importance of effective teamwork are initially ambivalent or negative.

Recommendation: Without a good understanding of the structure and content of the material being used for assessment and feedback, students are unlikely to get the most out of their training experiences. In future exercises of this kind, time should be devoted at the beginning of the event to explaining new or experimental material and approaches to enhance comprehension and motivation.

21. Both scores and comments are important for students learning – ratings show what went wrong, comments show how to fix it

The students involved in the CTT exercise expressed the view that both scores and comments were important as performance feedback in the Mentor feedback products. In the words of the students, scores tell them what is wrong, while comments tell them how to fix it. These comments reveal an implicit understanding of the difference between what has been termed outcome feedback and process feedback (e.g. Blickensderfer, Cannon-Bowers, & Salas, 1997). Outcome feedback provides information about the results of performance, and can be used to inform students of where changes in behaviour need to occur in future sessions. The Mentor system provides a large amount of outcome feedback in the form of the scores on objectives and measures. On the other hand, process feedback

provides information about specifically which aspects of behaviour should change and how students should go about making such changes. This typically involves much more detailed information than simple scores. The Mentor system provides for process feedback in the form of assessors' comments against objectives and measures. Both outcome and process feedback can be accessed by students either through the stoplight reports or the take-home handouts. While outcome feedback is often necessary, some researchers have claimed that it is usually not sufficient to achieve the best learning outcomes, particularly in team settings (e.g. Blickensderfer et al., 1997). It is therefore important that full use is made of the facility within the Mentor software to provide students both with information about where behaviour should change (outcome feedback, provided by scores), which specific behaviours should change and how (process feedback, provided by comments).

Recommendation: Both outcome and process feedback can be provided using the Mentor software. As both outcome and process feedback are important in their own way for achieving the best possible learning outcomes (particularly in team contexts) it is important that assessors are able to provide them both. This means that assessors must be (i) aware of this difference, and able to make observations relevant to each kind of feedback, and (ii) comfortable with the Mentor software interface prior to undertaking assessment duties, both in terms of selecting scores on measures and writing comments using handwriting recognition. Difficulty with any of these functions could dramatically reduce the quality of the feedback provided to the students in the form of Mentor feedback products.

22. Team-level assessment and feedback is important, but is currently under-emphasised

The Mentor software package is content free in the sense that it can (and indeed *must*) be populated by the user with roles, scenarios, events, objectives and measures pertinent to the training domain of interest. The user must go through the process of defining these characteristics of the training exercise or program – a process which can be time consuming and costly. The software merely provides a framework within which these elements may be organised and presented and data may be collected. As described in the introduction of this report, the Mentor package is primarily being put forward within the ADF as a tool for managing collective training. The FCC course exercise presented an opportunity to evaluate assessor and student reactions to both the Mentor software package itself and the more general concept of team-level performance assessment and feedback. It was the strong view of both assessors and students that training focused on enhancing teamwork was important, but under-emphasised in the current training curriculum.

Recommendation: Given the importance of knowledge, skills, and attitudes (KSAs) related to teamwork for enhancing the effectiveness of RAAF ABM teams, opportunities for enhancing team performance through principled approaches to collective training should be explored. Simply providing opportunities to practice individual tasks in a team context is not likely to yield the maximum benefit. Training focused on enhancing team performance should include objectives and scenarios aimed at stimulating critical team KSAs. The provision of timely and accurate feedback on teamwork-related behaviours

must also be a key consideration. Given the strengths of the Mentor software in managing these aspects of training it is potentially a very useful tool for this purpose.

23. The team dimensions were generally easy to understand

The research literature on teamwork and team training is replete with taxonomies purporting to describe the critical dimensions of teamwork (see Lenne, 2003 for a review). This can lead to confusion, as the literature provides many different ways in which one can conceptualise teamwork and effective team performance. However, close inspection reveals that there is a substantial amount of commonality between many different teamwork taxonomies. A taxonomy of critical teamwork dimensions which appears to capture the important determinants of team effectiveness in a relatively simple factor structure, and is well-founded in empirical research is that reported by Smith-Jentsch, Johnston, and Payne (1998). This taxonomy arose from the US Navy sponsored TADMUS (Tactical Decision Making Under Stress) program. According to this view, critical teamwork behaviours can be grouped into four dimensions, being related to: (i) information exchange, (ii) communication, (iii) supporting behaviour, and (iv) initiative/leadership. These dimensions formed the basis of the teamwork objectives and measures that were entered into Mentor and used for the ABM team training exercise. An additional dimension, named 'team coordination', was made available to the assessors for this exercise. The inclusion of this dimension facilitated assessment and feedback of such factors as team members' awareness of each other's tasks, the distribution of workload within the team and the consistency of actions with plans.

Assessors and students involved in the CTT exercise reported that they found this particular way of conceptualising teamwork to be relatively easy to comprehend and work with (but see Point 20 above). This is an important point, since the ease with which participants can understand teamwork concepts could be expected to affect both the way in which assessors assign ratings and comments and the way in which students conceptualise their performance and plan behavioural changes based on feedback.

Recommendation: The ease with which participants can comprehend teamwork concepts is likely to be an important determinant of team training effectiveness. The approach taken for the exercise reported here was assessed as relatively easy to understand by both assessors and students. This approach is based on sound empirical evidence. Therefore, this taxonomy of teamwork dimensions should be considered when defining objectives and measures for future team training exercises.

24. Team assessment is more appropriate as a learning activity than an assessment activity

While the assessors viewed KSAs related to effective teamwork as important, they felt that evaluation of team performance, as opposed to individual performance, was more appropriate as a learning activity than as an assessment activity. This is an important issue, because assessment at a collective level, such as that undertaken here, can be conceptualised as a means to qualify or certify the 'readiness' of organisational units (hence the proposal to use Mentor as the basis for an Air Warfare Assessment and Readiness Evaluation System; AWARES).

Recommendation: The matter of how collective assessment is used within the RAAF is one which must be addressed through consideration of organisational goals and values. While it is outside the scope of this report to attempt to resolve such issues, it is important to note the potential value of collective assessment, if supported and conducted appropriately, for monitoring the readiness status of organisational units.

25. Team training should be considered an adjunct to individual training
While students and assessors agreed that team training was important for promoting the effectiveness of their organisation, they also emphasised the importance, and prerequisite nature of, individual professional mastery. That is to say, the participants interviewed for this report believed that team training should be seen as a way to enhance the integration of contributions to collective performance from *already-highly-expert* individual operators. The individual versus collective task distinction can be thought of as a dimension of workplace complexity. This view is therefore consistent with a graded approach to training for complex work environments, which consists of introducing complexity in a measured fashion over the course of time.

Recommendation: By definition (e.g. Paris, Salas & Cannon-Bowers, 2000) effective teamwork involves the coordination of inputs from two or more people in working towards a common goal. The interpersonal coordination required for good teamwork involves a number of cognitive and social skills. It is likely that such skills are best trained when students have sufficient cognitive resources to devote to them; as when a degree of expertise on important elements of individual tasks has already been achieved. This supports a graded, crawl-walk-run approach to integrating individual and team training.

3. Performance Change: Quantitative Analysis

The interview outcomes presented above provide information on usability and user acceptance to guide future system development. However, an equally important aspect of training system evaluation is performance change. Data from the two expert assessors were gathered using Mentor for the purpose of comparing the frequencies of SAT, MARG and UNSAT ratings given across the sessions. The aim was to use the Mentor data to examine performance change for (i) the test team (TT) who were assessed and received Mentor team-level feedback products (i.e. team debrief structured around Mentor stoplight reports and handouts), and (ii) the control team (CT) who were assessed using the Mentor system to provide comparison data, but who were not provided with feedback structured around Mentor feedback products. It was expected that a comparison of performance change between the two teams could provide evidence regarding the impact of receiving such feedback. The raw frequencies of ratings in each category for the two teams are summarised in Table 4 below.

Table 4. Frequency of ratings allocated to each category (SAT, MARG, UNSAT) during each session of the exercise

		SAT	MARG	UNSAT	Total (Unrated)
Test Team	First Session	24	10	0	34 (49)
	Mid-Ex Session	49	4	0	53 (39)
	Last Session	28	0	0	28 (35)
Control Team	First Session	36	6	0	42 (34)
	Last Session	38	1	0	39 (24)

As can be seen from the right-hand column of Table 4, the assessors did not rate all of the items that were available to them during any of the sessions. This situation most likely arose due to the fact that the Mentor objectives and measures were developed separately from the exercise scenario events. This led to a less direct mapping between the scenario events, objectives and measures than would ideally be the case. Because of this, a liberal strategy was adopted when considering which measures to link to which serial events: All measures which the assessors could conceivably find useful for a given serial were included. Both assessors rated fewer items in their last session using the Mentor DET than they did in their first session. This pattern of results is inconsistent with a learning effect in which assessors initially found the DET difficult to manage, but gradually developed a level of proficiency. Given that the sessions gradually became more complex over the course of the exercise, this could be an effect of assessor workload – with the increased cognitive effort required to enter data into the Mentor DET (see Theme 7 in Section 2), assessors may have been unable to rate many items and provide comments when scenario events involved many aircraft or occurred in very rapid succession.

While the quantitative data are useful in understanding the way in which the assessors used the Mentor tool, there are problems with their use for evaluating the specific impact of the Mentor system on team performance. In particular, the comparison of performance between the two teams was significantly compromised by aspects of the exercise design which were imposed due to technical, scheduling, and personnel availability issues.

A number of factors contributed to making the comparison between ABM teams in this exercise problematic. First, due to availability and scheduling, the TT and the CT were assessed by different instructors, neither of whom were blind to the conditions of the evaluation. While these instructors were both SACTU experts, it is possible that differences observed between the two teams could be due to expectancy effects or differences between the rating tendencies or predispositions of the instructors. While it is difficult to conceive of how the assessors might be made blind to conditions in evaluations such as this (e.g. the assessor must lead the stoplight AAR), the use of different assessors for different teams should be avoided. Second, as noted in Section 1, due to the availability of personnel, one member of the CT was also a member of the TT. It is possible that exposure to the Mentor feedback products led to behaviour changes which affected the performance of both teams of which this individual was a member. Third, because of

technical issues with the simulator, the CT had their first session (Day 1 AM) cancelled, while the TT did not. Because of this, the TT actually had one more session during the exercise than the CT. It is possible that any performance difference observed between the teams could simply be due to the TT receiving more training than the CT. And fourth, due to scheduling adjustments, the first Mentor team-level debrief was very rushed, taking only a few minutes to complete. It is likely that this reduced the effect that the feedback had on the performance of the members of the TT during subsequent simulator sessions. In the presence of these confounding factors it was not possible to identify with certainty the effect of receiving feedback via the Mentor products. This outcome and its implications for future training system evaluations are discussed further in Section 4.

While the particular effects of Mentor on performance in this exercise could not be determined, the presence of the confounding factors did not preclude an examination of overall performance change during the exercise. That is, data from the two assessors could still be used to determine whether the teams who took part in the simulation exercise benefited from their involvement as demonstrated by a performance improvement over the course of the two days.

To examine this question, the raw frequencies of SAT, MARG and UNSAT ratings from both teams were combined into a single data set and considered across the three sessions for which data was collected (i.e. Day 1 AM, Day 2 AM, and Day 2 PM). To achieve a useful comparison between sessions containing different numbers of available and rated measures, the frequencies displayed in Table 4 were expressed in terms of the proportion of assigned ratings (i.e. not including unrated measures) that were rated as SAT. These proportions are presented in Figure 4 below.

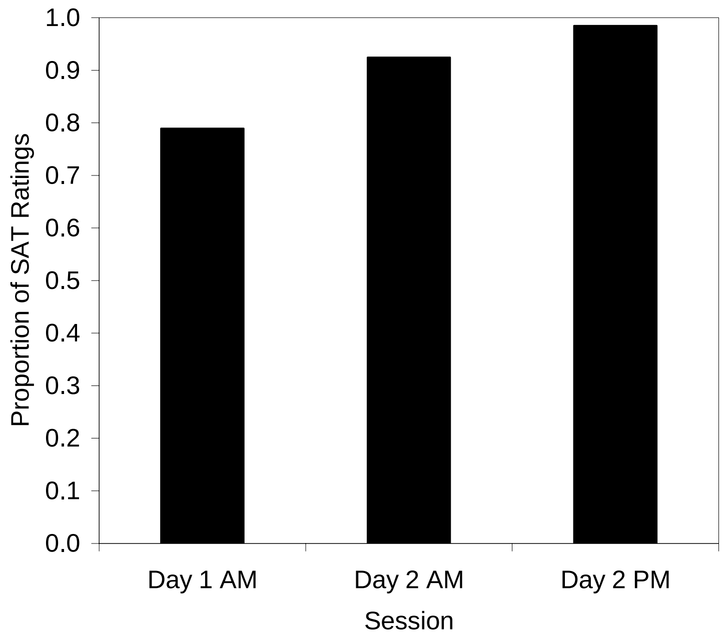


Figure 4. Proportion of SAT ratings given to the test and control teams (combined) in the sessions for which Mentor data was collected

Note that as shown in Table 4, no UNSAT ratings were given by the instructors during any of the sessions. This means that no proportions are displayed in Figure 4 for UNSAT ratings. MARG ratings are also not shown, as they are redundant with the SAT ratings (i.e. for any given session MARG = 1 – SAT). As can be seen from the chart presented in Figure 4, assessor ratings of team performance increased over the two days of the exercise.

The scenario for this exercise was designed such that hostilities escalated from session to session. That is, each session was specifically designed by the exercise coordinator to be more demanding on the ABM team than those preceding it. If it is assumed that the exercise coordinator was indeed successful in designing the sessions as such, the results presented above provide evidence in favour of a benefit for ABM teams from this kind of team training. In terms of the objectives and measures used on this occasion (see Appendix A), the ABM teams appeared to improve their level of performance in the face of ever more demanding circumstances over the course of the two days. This apparent improvement was observed across the whole range of objectives and measures, such that during the last session of Day 2 only a single MARG rating was assigned out of a combined total of 67 rated measures[5].

4. Summary and Conclusions

All students and instructors involved in this evaluation considered collective training, assessment and feedback to be important activities for improving the effectiveness of RAAF ABM teams. However, they also lamented the fact that the opportunities for such training come about relatively infrequently when compared to individual instruction. The evidence presented here suggests that training as a team does lead to at least short term performance improvements on behavioural observation measures related to ABM team tasks and important teamwork dimensions.

The qualitative data reported in Section 2 revealed positive trainee and assessor reactions to the Mentor system. However, problems arising from the conduct of the evaluation rendered the quantitative data reported in Section 3 of limited value for determining the impact of the Mentor system on trainee performance. Due to the presence of significant confounding factors, it was not clear from the data collected during this evaluation whether the use of the Mentor system led to greater improvements in performance over the course of the exercise than would otherwise have been the case. It is important to note why these confounding factors came about. The shortcomings of this evaluation arose largely because the simulation exercise reported here was part of a course aimed, first and foremost, at training and qualifying students. While the evaluation of the use of the Mentor system for collective training events was the primary goal of the DSTO human factors team, this evaluation and the maintenance of the experimental conditions required to draw valid conclusions from it, were not the primary objectives of the event as a whole.

[5] The single MARG rating assigned during the afternoon sessions of Day 2 was assigned to the CT on the item "Efficient yet effective airborne posture continually maintained". During the afternoon session of Day 2 all other rated measures for both teams were rated SAT.

The quality of the experimental design for the evaluation was, on a number of occasions, balanced against practical considerations such as scheduling and the availability of personnel. In virtually all cases, sacrifices in experimental design were required. These sacrifices meant that it was not possible to demonstrate a compelling case for the effectiveness of the Mentor tool in terms of enhancing team performance.

The incompatibility evidenced here between the practicalities of running training events and the requirements of rigorous training research represents a challenge for the future of training in the ADF. The adoption of new training strategies, or alternatively, the maintenance of legacy approaches must be based on solid empirical research if the best outcomes are to be achieved. While it is relatively straightforward to collect 'reactions' data (i.e. how participants felt about the training they received or administered) under most circumstances, Kirkpatrick's widely applied model of training system evaluation asserts that it is necessary to demonstrate more than positive reactions. Behaviour change during training events, transfer of training to on-the-job performance, and the fit between organisational goals and training goals must also be examined in order to ensure that training strategies deliver maximum benefits. While some of these points can be addressed in the laboratory or other contexts, some (e.g. the issue of transfer of training) can only be addressed by conducting well-designed research in operational contexts. This will only be possible if appropriate experimental design is given high priority during the planning and conduct of events within which such research activities are to take place. In order to reach valid conclusions, empirical investigations must be designed and executed carefully. Departures from experimental design applied during run time will usually have significant impacts on the quality of the outcomes that are achieved. The confounding factors described above and the attendant shortcomings of the quantitative, performance-based findings reported here are a case in point.

In order to achieve greater confidence regarding the outcomes of training research activities in the future, more controlled research environments are required. However, this does not mean that such evaluations can, or should, only be conducted in the laboratory. Close cooperation is required between researchers, exercise managers, run-time controllers, assessors, and trainees to ensure that rigorous training research can take place outside of the laboratory. This will maximise the validity and generalisability of research findings as well as the justifiability of decisions regarding training strategies, tools, and techniques.

In summary, while there were limits to the conclusions that could be made here, it is clear from this evaluation that the Mentor system performs well in facilitating planning, assessment, and provision of timely feedback in team training contexts. While some shortcomings have been identified, Mentor has many useful features and good user acceptance. Further evaluations of the use of Mentor should take place as the system develops and matures in order to take full advantage of the support it provides to collective training within the RAAF and the ADF more broadly. However, clearer evidence regarding the particular effects of this and other systems on trainee performance demands investigation in more controlled environments, potentially over longer periods of time, and should include examination of transfer of training. In order to achieve this, the

requirements of effective training events and those of valid research activities must be reconciled.

5. References

Blickensderfer, E., Cannon-Bowers, J.A., & Salas, E. (1997). Theoretical bases for team self-correction. *Advances in Interdisciplinary Studies of Work Teams, 4,* 249-279.

Clark, P., Lenne, M., Robbie, A., Ross, P., Ryan, P., & Zalcman, L. (2003). *Coalition Readiness Management System (CReaMS) Phase Three: Virtual Coalition Readiness*. DSTO Technical Report. DSTO-TR-1535

Kirkpatrick, D.L. (1987). Evaluation. In R.L. Craig (Ed.) *Training and development Handbook: A Guide to Human Resource Development* (2nd Ed). New York: McGraw-Hill.

Lenne, M. (2003). *A Review of Critical Human Factors Issues for Aviation Team Training.* DSTO Technical Report. DSTO-TR-1466.

McCarthy, A., Kingston, G., Johns, K., Gori, R., Main, P. & Kruzins, E. (2003). *Joint Warfare Capability Assessment - Final Report: Australian Joint Essential Tasks Volume 1*. DSTO Client Report. DSTO-CR-0293.

Matthews, M.L. & Lamoureaux, T.M. (2003). *Development of Generic Aircrew Measures of Performance for Distributed Mission Training*. DRDC Report CR-2003-060.

Paris, C.R., Salas, E., & Cannon-Bowers, J.A. (2000). Teamwork in multi-person systems: A review and analysis. *Ergonomics*, 43(8), 1052-1075.

Smith-Jentsch, K.A., Johnston, J.H., & Payne, S.C. (1998). Measuring team-related expertise in complex environments. In J. Cannon-Bowers & E.Salas (Eds.) *Making Decisions Under Stress: Implications for Individual and Team Training* (pp. 61-87). Washington, DC: APA Press.

Vicente, K.J. (1999). *Cognitive Work Analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wells, G. L., & Loftus, E. F. (2003). Eyewitness memory for people and events. In A. M. Goldstein (Ed.), *Handbook of psychology: Forensic psychology*, Vol. 11. (pp. 149-160). Hoboken, NJ: John Wiley & Sons.

Wickens, C.D., & Hollands, J.G. (2000). *Engineering Psychology and Human Performance* (3rd Ed). Upper Saddle River, NJ: Prentice Hall.

Appendix A: Mentor Objectives and Measures

Table A1. Objective Categories (ASJETS), Objectives and Measures used to populate the Mentor database for the CTT exercise reported here

Objective Category (ASJETS)	Objective	Measures
Organise Command and control	External liaison	External verbal comms clear, concise and in correct format
		External comms are performed using correct means
	Internal Liaison	Internal verbal comms clear, concise and in correct format
		Internal comms used correct format
Implement ATO/ACO	Aircraft safety	Separation standards adhered to
		Non Participating aircraft detected and addressed in timely manner
		Appropriate actions taken on separation breakdown
		Airspace breaches pre-emptively avoided
Airspace Management	ADIZ Procedures	ADIZ Procedures are enforced
		Airborne requests for transit authorised properly
		All aircraft entering ADIZ are identified in a timely manner
		Challenge procedures issued
		Unauthorised aircraft intercepted
		Verbal warnings issued over GUARD
	Maintain Safety of Flight	Separation standards enforced
		Mercy flight clearances provided
		Emergency procedures followed appropriately
		Airspace managed efficiently and IAW procedures
		Aircraft maintained within allocated space
		Accurate and concise clearances issued
		Aircraft recovery/handoff/transit co-ordinated in timely manner
Maintain Situational Awareness	Maintain Team SA	Air contacts classified IAW procedures

Maintain Situational Awareness (cont)	Maintain Team SA (cont)	RAP maintained in a timely manner
		Appropriate radio frequencies constantly monitored
		Sensors managed to ensure most effective surveillance product
		Asset tote board maintained
		System status maintained
	System Degrade	Correct degraded system procedures applied in a timely manner
		Disruptions to system performance handled seamlessly
Conduct Defensive Counter Air	Set & Maintain Posture	Efficient yet effective ground ALERT posture maintained
		Efficient yet effective Airborne posture continually maintained
		Sufficient Defence in Depth
		Awareness of asset status continually maintained
		Effective low-level sanitisation
	Tactical Employment	Pre-emptive Inter-FEZ Co-ordination
		Authentication procedures enforced
		Intercepts conducted IAW Briefed procedures/SIs
		Weapons employed efficiently
Implement ROE and Request Changes	Apply ROE	ROE valid for all engagements
		ROE Matrix satisfied on timeline
	Modify ROE	Timely extensions to ROE requested
Collect Info on Enemy ORBAT & Targets	Report Activity	Air Raids reported
		Enemy tactics reported
Debrief/Review Mission	Brief	Mission requirements clearly understood
		Commanders intent clearly understood
		Ambiguous brief elements clarified where appropriate
Demonstrate Effective Teamwork	Communication	Proper phraseology used
		Standard reporting procedures followed
		Information spoken/delivered clearly and succinctly

Demonstrate Effective Teamwork (cont)	Communication (cont)	Reports from team mates acknowledged
		Proper phraseology used
	Information Exchange	All sources of information used effectively
		Information passed without having to be asked
		Big Picture "updates as appropriate"
		All relevant information shared among team members
	Initiative/ Leadership	Guidance and suggestions provided to team mates as appropriate
		Tasks allocated according to appropriate priorities
	Supporting Behaviour	Errors promptly identified and corrected
		Back-up or assistance provided when needed
		Back-up or assistance requested when needed
	Team Co-ordination	Team members maintain awareness of others' tasks
		Team members facilitate the performance of others' tasks
		Workload distributed appropriately
		Planned actions implemented appropriately

Note: The Mentor stoplight report templates available at the time of this exercise required a four-level objective hierarchy in order to function. Because of this, the top level of the objective hierarchy detailed in the table above had to be repeated in the stoplight reports. This had only a small impact, in that the assessor was required to drill down two levels, rather than one, in order to access the 'Objective' level of the report.

DISTRIBUTION LIST

“As per the Research Library’s *Policy on electronic distribution of official series reports* (http://web-vic.dsto.defence.gov.au/workareas/library/aboutrl/roles&policies/mission.htm) Unclassified (both Public Release and Limited), xxx-in-confidence and Restricted reports and their document data sheets will be sent by email through DRN to all recipients with Australian defence email accounts who are on the distribution list apart from the author(s) and the task sponsor(s). Other addressees and Libraries and Archives will also receive hardcopies.”

Evaluation of the 'Mentor' Assessment and Feedback System for Air Battle Management Team Training

Christopher Best and Eleanore Burchat

AUSTRALIA

DEFENCE ORGANISATION	**No. of copies**
Task Sponsor	
DGAD	1 Printed
S&T Program	
Chief Defence Scientist	1
Deputy Chief Defence Scientist Policy	1
AS Science Corporate Management	1
Director General Science Policy Development	1
Counsellor Defence Science, London	Doc Data Sheet
Counsellor Defence Science, Washington	Doc Data Sheet
Scientific Adviser to MRDC, Thailand	Doc Data Sheet
Scientific Adviser Joint	1
Navy Scientific Adviser	Doc Data Sht & Dist List
Scientific Adviser – Army	Doc Data Sht & Dist List
Air Force Scientific Adviser	1
Scientific Adviser to the DMO	Doc Data Sht & Dist List
Deputy Chief Defence Scientist Platform and Human Systems	1
Chief of Air Operations Division	1
Research Leader, Crew Environments and Training	1 Printed
Head of Training Research	1 Printed
Task Manager – Christopher Best	1
Author: Christopher Best	1 Printed
Author: Eleanore Burchat	1 Printed
Jon Blacklock	1 Printed
Michael Skinner	1 Printed
Peter Ryan	1 Printed
Peter Clark	1 Printed
David McIlroy	1 Printed
George Galanis, Land Operations Division	1

DSTO Library and Archives	
Library Fishermans Bend	1 Printed
Library Edinburgh	1 Printed
Defence Archives	1 Printed
Capability Development Group	
Director General Maritime Development	Doc Data Sheet
Director General Land Development	Doc Data Sheet
Director General Capability and Plans	Doc Data Sheet
Assistant Secretary Investment Analysis	Doc Data Sheet
Director Capability Plans and Programming	Doc Data Sheet
Chief Information Officer Group	
Head Information Capability Management Division	Doc Data Sheet
Director General Australian Defence Simulation Office	Doc Data Sheet
AS Information Strategy and Futures	Doc Data Sheet
Director General Information Services	Doc Data Sheet
Strategy Group	
Assistant Secretary Strategic Planning	Doc Data Sheet
Assistant Secretary International and Domestic Security Policy	Doc Data Sheet
Navy	
Maritime Operational Analysis Centre, Building 89/90 Garden Island Sydney NSW Deputy Director (Operations) Deputy Director (Analysis)	Doc Data Sht & Dist List
Director General Navy Capability, Performance and Plans, Navy Headquarters	Doc Data Sheet
Director General Navy Strategic Policy and Futures, Navy Headquarters	Doc Data Sheet
Air Force	
SO (Science) - Headquarters Air Combat Group, RAAF Base, Williamtown NSW 2314	Doc Data Sht & Exec Summary
Staff Officer Science Surveillance and Response Group	Doc Data Sht & Exec Summary
Army	
Australian National Coordination Officer ABCA (AS NCO ABCA), Land Warfare Development Sector, Puckapunyal	Doc Data Sheet
J86 (TCS GROUP), DJFHQ	Doc Data Sheet
SO (Science) - Land Headquarters (LHQ), Victoria Barracks NSW	Doc Data Sht & Exec Summary
SO (Science) - Special Operations Command (SOCOMD), R5-SB-15, Russell Offices Canberra	Doc Data Sht & Exec Summary & Dist List

SO (Science), Deployable Joint Force Headquarters (DJFHQ) (L), Enoggera QLD	Doc Data Sheet
Joint Operations Command	
Director General Joint Operations	Doc Data Sheet
Chief of Staff Headquarters Joint Operations Command	Doc Data Sheet
Commandant ADF Warfare Centre	Doc Data Sheet
Director General Strategic Logistics	Doc Data Sheet
Intelligence and Security Group	
AS Concepts, Capability and Resources	1
DGSTA, Defence Intelligence Organisation	1
Manager, Information Centre, Defence Intelligence Organisation	1
Director Advanced Capabilities	Doc Data Sheet
Defence Materiel Organisation	
Deputy CEO	Doc Data Sheet
Head Aerospace Systems Division	Doc Data Sheet
Head Maritime Systems Division	Doc Data Sheet
Program Manager Air Warfare Destroyer	Doc Data Sheet
Guided Weapon & Explosive Ordnance Branch (GWEO)	Doc Data Sheet
CDR Joint Logistics Command	Doc Data Sheet

OTHER ORGANISATIONS

National Library of Australia	1
NASA (Canberra)	1

UNIVERSITIES AND COLLEGES

Australian Defence Force Academy	
Library	1
Head of Aerospace and Mechanical Engineering	1
Hargrave Library, Monash University	Doc Data Sheet

OUTSIDE AUSTRALIA

INTERNATIONAL DEFENCE INFORMATION CENTRES

US Defense Technical Information Center	1
UK Dstl Knowledge Services	1
Canada Defence Research Directorate R&D Knowledge & Information Management (DRDKIM)	1
NZ Defence Information Centre	1

ABSTRACTING AND INFORMATION ORGANISATIONS

Library, Chemical Abstracts Reference Service	1
Engineering Societies Library, US	1
Materials Information, Cambridge Scientific Abstracts, US	1
Documents Librarian, The Center for Research Libraries, US	1

INFORMATION EXCHANGE AGREEMENT PARTNERS

National Aerospace Laboratory, Japan	1
National Aerospace Laboratory, Netherlands	1

SPARES	5 Printed

Total number of copies: 45 Printed: 18 PDF: 27

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION
DOCUMENT CONTROL DATA

1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)

2. TITLE Evaluation of the 'Mentor' Assessment and Feedback System for Air Battle Management Team Training	3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION) Document (U) Title (U) Abstract (U)
4. AUTHOR(S) Christopher Best and Eleanore Burchat	5. CORPORATE AUTHOR DSTO Defence Science and Technology Organisation PO Box 1500 Edinburgh South Australia 5111 Australia

6a. DSTO NUMBER	6b. AR NUMBER	6c. TYPE OF REPORT	7. DOCUMENT DATE
DSTO-TR-1942	AR-013-795	Technical Report	November 2006

8. FILE NUMBER	9. TASK NUMBER	10. TASK SPONSOR	11. NO. OF PAGES	12. NO. OF REFERENCES
2006/1114986/1	AIR 04/236	DGAD	38	11

13. URL on the World Wide Web	14. RELEASE AUTHORITY
http://www.dsto.defence.gov.au/corporate/reports/DSTO-TR-1942.pdf	Chief, Air Operations Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for public release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111

16. DELIBERATE ANNOUNCEMENT

No Limitations

17. CITATION IN OTHER DOCUMENTS Yes

18. DSTO RESEARCH LIBRARY THESAURUS http://web-vic.dsto.defence.gov.au/workareas/library/resources/dsto_thesaurus.htm

Training systems, Australian Defence Force, team performance, teamwork

19. ABSTRACT

The Mentor software package (Calytrix Technologies, Perth, Western Australia) is gaining popularity within the Australian Defence Force (ADF) as a means by which to manage training objectives, collect performance data and provide feedback for collective training. While the Navy has led the way in the application of this tool, it is now being put forward as an important component of an Air Warfare Assessment and Readiness Evaluation System (AWARES) for the RAAF as well as being included in the suite of tools to be used for exercises involving the Joint Combined Training Centre (JCTC). This report contains an account of an evaluation of the Mentor system and its use to provide performance assessment and feedback during a RAAF Air Battle Management team training event.